**Exercise 11: Using Bioinformatics to study phylogenetic relationships;**

Are polar bears closely related to American brown bears? Do you think the giant panda is more closely related to the American brown bear or the polar bear? Is the giant panda a bear at all? What about the red panda – is it a giant panda relative? What does the evolutionary history of bears look like? In this activity you will use real gene sequences to figure out which animals are more closely related. You will create a family tree for bears and help to solve the mystery of bear history!

**TABLE 1: Accession numbers and scientific names different species**

|  |  |  |
| --- | --- | --- |
| **Organism** | **Scientific Name** | **Accession Number** |
| American black bear |  | Y08520.1 |
| American brown bear |  | L21889.1 |
| Spectacled bear |  | L21883.1 |
| Asiatic bear |  | L21890.1 |
| Polar bear |  | L22164.1 |
| Giant panda |  | Y08521.1 |
| Red Panda |  | S80939.1 |
| Northern raccoon |  | U78345.1 |
| Leopard seal |  | AY377288.1 |

1. Go to [**www.ncbi.nlm.nih.gov**](http://www.ncbi.nlm.nih.gov). Click the arrow beside “all databases” and select **Nucleotide**. Enter the **accession number** for the American Black bear (see Table 1 for number) in the search space.
2. You will see lots of information. Near the top you will see the definition listed as: *U. americanus* mitochondrial 12S rRNA gene. *U. americanus* is the scientific name for the American black bear. The gene is part of the mitochondrial DNA and makes up part of each ribosome. Write the scientific name for the American black bear in Table 1.
3. Click **FASTA** at the top of the page. You will then see the sequence of DNA that makes up this 12S rRNA gene.
4. On the right, towards the top, is a drop-down arrow next to “Send”. Click the arrow and choose **Clipboard** as your destination. Then click “**Add to Clipboard**.”
5. Continue by searching each animal’s accession number, clicking on the FASTA, and sending the sequences to the clipboard.
6. After you have added all 9 FASTA sequences to the clipboard, click on the “**9 items**” next to clipboard to open your clipboard.
7. **Select each animal** by checking the box.
8. After all nine have been selected, click the “**Send to**” in the upper right hand corner.
9. Chose “**File**” and under Format select “**FASTA**”.
10. Chose “**Organism Name**” under “**Sort By**”.
11. Click “**Create File**”
12. Your file will be created with the following name: **sequence (1).fasta**, which will be automatically downloaded. You will be able to find this file in your download files.

You will now take these sequences for the 12S rRNA gene and align them using an online website. The website uses algorisms (math) to find areas of similarity across the 9 organisms. This requires significant computation so this is always done via computers.

1. Now you will go to [**http://www.genome.jp/tools/clustalw**](http://www.genome.jp/tools/clustalw).
2. Chose “**CLUSTAL**” for the output format.
3. Check “**SLOW/ACCURATE**” for the pairwise alignment.
4. Check “**DNA**” for Enter your sequences.
5. You will upload the sequence file you just made by choosing “**Choose File**”. Navigate to find the file you downloaded called **sequence (1).fasta**.
6. Click “**Execute Multiple Alignment**”.
7. The program will align the sequences to see the results.

***Interpreting your results***

1. At the top you will see the 9 sequences with their accession numbers and the number of base pairs (bp) for the sequence. The program numbered your organisms 1 – 9. In the table below, write the common name beside each number. You must look back at the accession numbers to match the right name to the correct number.

|  |  |
| --- | --- |
| **Organism Number** | **Organism Name** |
| **1** |  |
| **2** |  |
| **3** |  |
| **4** |  |
| **5** |  |
| **6** |  |
| **7** |  |
| **8** |  |
| **9** |  |

1. Now scroll down slightly and you will see alignment scores like below:

Sequences (1:2) Aligned. Score: 75.

1. This is just an example and will not match your data! This is the score between organism 1 and organism 2. The **higher** the **alignment score** between two organisms, the **more similar** their **DNA** is for this gene.
2. Scroll down farther and look at the sequence alignments; you will see nucleotide sequences (the A’s, G’s, T’s, and C’s). The symbol “–” indicates that there is no information for this location in the sequence. In other words the base (letter) is not known. Because some sequences are longer than others there may be many sections where information is missing. The symbol “**\***”indicates a location where all nine organisms have the exact same nucleotide, meaning the base is conserved. If a nucleotide is conserved then that same nucleotide should be in the common ancestor.

***QUESTIONS ABOUT THE ALIGNMENT***

1. Which two species have the highest alignment scores? You will need to check the accession numbers so you know which organism is which.
2. Which two species have the lowest alignment scores?
3. Which group has the highest score with the red panda?
4. Which group has the second highest score with the red panda?
5. Which group has the lowest score with the red panda?
6. What does it mean if two groups have a high alignment score?
7. Now you will use this website to build a phylogenic tree. This is like a family tree based on genetics and evolution. The website takes the data that you have looked at and draws a tree, which is more visually appealing.
8. At the top of the page (above the alignment data), select “Rooted phylogenetic tree with branch length (UPGMA)” from the “Select tree menu” drop-down menu.
9. Click on “**Exec**”.
10. You will see the phylogenic tree. What does this tell you about the relationships among bear species?

**Exercise 12 “Extra Credit”: Can you figure out what species these proteins belong to? Can you create a phylogenetic tree? Can you find the DNA sequences for these proteins?**

Species One: Beta-hemoglobin Amino Acid Sequence

1. mvhltpeeks avtalwgkvn vdevggealg rllvvypwtq rffesfgdls tpdavmgnpk

61 vkahgkkvlg afsdglahld nlkgtfatls elhcdklhvd penfrllgnv lvcvlahhfg

121 keftppvqaa yqkvvagvan alahkyh

Species Two: Beta-hemoglobin Amino Acid Sequence

1 mvhltpeeks avtalwgkvn vdevggealg rllvvypwtq rffesfgdls tpdavmgnpk

61 vkahgkkvlg afsdglahld nlkgtfatls elhcdklhvd penfrllgnv lvcvlahhfg

121 keftppvqaa yqkvvagvan alahkyh

Species Three: Beta-hemoglobin Amino Acid Sequence

1 mvhltpeeks avtalwgkvn vdevggealg rllvvypwtq rffesfgdls tpdavmgnpk

61 vkahgkkvlg afsdglahld nlkgtfatls elhcdklhvd penfkllgnv lvcvlahhfg

121 keftppvqaa yqkvvagvan alahkyh

Species Four: Beta-hemoglobin Amino Acid Sequence

1 vhltpeeksa vtalwgkvnv devggealgr llvvypwtqr ffesfgdlst pdavmgnpkv

61 kahgkkvlga fsdglahldn lkgtfaqlse lhcdklhvdp enfrllgnvl vcvlahhfgk

121 eftpqvqaay qkvvagvana lahkyh

Species Five: Beta-hemoglobin Amino Acid Sequence

1 vhltpeekna vttlwgkvnv devggealgr llvvypwtqr ffdsfgdlss paavmgnpkv

61 kahgkkvlga fsdglnhldn lkgtfaqlse lhcdklhvdp enfkllgnvl vcvlahhfgk

121 eftpqvqaay qkvvagvana lahkyh

Species Six: Beta-hemoglobin Amino Acid Sequence

1 vhltgeeksa vttlwgkvnv eevggealgr llvvypwtqr ffesfgdlss pdavmgnpkv

61 kahgkkvlga fsdglahldn lkgtfaqlse lhcnklhvdp enfrllgnvl vcvlahhfgk

121 eftpqvqaay qkvvagvana lahkyh

Species Seven: Beta-hemoglobin Amino Acid Sequence

1 vhltgeeksa vttlwgkvnv eevggealgr llvvypwtqr ffesfgdlss pdavmgnpkv

61 kahgkkvlga fsdglahldn lkgtfaqlse lhcnklhvdp enfrllgnvl vcvlahhfgk

121 eftpqvqaay qkvvagvana lahkyh